RN/05/25
November 2, 2005

Research
Note

The Coalition Peering Domain: A New Entity on the Routing Landscape

Manish Lad

Telephone: +44 (0)20 7679 3666
Fax: +44 (0)20 7679 1397
Electronic Mail: m.lad@cs.ucl.ac.uk
URL: http://www.cs.ucl.ac.uk/staff/m.lad/

Abstract

This report details the investigations undertaken to date in the general area of community-area networking. It proposes that existing networking mechanisms are not suited sufficiently to enable a newly emerging class of community-area network; one where individuals connect together directly their home and personal-area networks, to form a local neighbourhood *mesh* or *community network*. A background and motivation for further research into this problem space is provided and a work-plan is outlined to better enable this newly emerging class of community-area network. The work-plan includes proposals to highlight also the wider-reaching implications of the thesis of this work beyond the neighbourhood context.

Keywords

Coalition-Based Community Networking

UCL

Department of Computer Science
University College London
Gower Street
London WC1E 6BT, UK

Contents

1 Area of Research

A new class of community-area networks [7, 4, 3, 9] are emerging. They involve individuals connecting together directly their home and personal-area networks, forming a local neighbourhood *mesh* or *community network*. However, such interconnection, or peering, is usually carried out by individuals on an ad hoc basis and there are no defined rules or protocols to facilitate the formation of these community-area networks. Each peering agreement between pairs of community members is unique.

This work aims to investigate how this newly emerging class of community-area networks may benefit from the structuring of the ad hoc peering, in a way that encourages collaboration while maintaining local control.

1.1 Background

Local-area networking capabilities have improved greatly in recent years. This allows users to inter-connect easily multiple machines or devices to utilise more efficiently and flexibly both their local resources and their access to the wide-area. Use of both wired Ethernet and wireless IEEE 802.11 standards have increased, while costs of both types have fallen significantly in recent years; manufacturers now integrate them into their equipment (e.g. laptops, desktops and ADSL gateways). Many consumer operating system platforms also provide improved networking support. They enable very simple local network set-up in a plug-and-play manner by configuring 'connection sharing' automatically through a combination of Network Address Translation (NAT) and automatic address allocation using the Dynamic Host Configuration Protocol (DHCP).

Advances in Internet and wide-area access technologies also have had a great impact on connectivity both for the home user and for the mobile user. The data rates of wired Internet access technologies (e.g. ADSL) have increased substantially, improving the speed at which users of edge networks are able to download and access content, and fueling the demand for various types of (multimedia) content that often involve large downloads of data. The recent popularity of peer-to-peer applications (especially peer-to-peer file-sharing applications) means that users of edge networks may now be interested in high upstream capacity on access links as well as high downstream capacity.

However, it is only very recently that levels of subscription to technologies like ADSL have increased to their current levels in some parts of the world. Meanwhile, in many parts of the world, there still remains a large proportion of home users for whom primary connection to the Internet is through older, slower technologies: mainly analogue modems but some ISDN. There still remains also a large number of users for whom the only connectivity in a mobile environment is through GSM. Thus we have a growing situation where there is very good local-area connectivity supported by both wired and wireless networking technologies, but where wide-area data rates (both wired and wireless) are improving relatively slowly and vary greatly leaving a large disparity.

1.2 Motivation

> "...for the Internet, much of the creative energy is at or near the edge of the network. It is at the edge that most applications are created. It is at the edge that most devices are connected. It is at the edge where we usually see the development of new networking technologies. It is at the edge of the network where the economic conditions most favour innovation, as the barrier to entry (for applications, devices, and networking technologies) is typically lower at the edge. And, at a fundamental level, the purpose of an Internet is to hook computers and similar "things" together, and we connect new "things" at the edge.
>
> So, if we want to think about where networking might be in 10 or 15 years, it behoves us to look at (r)evolution at the edge."
>
> — Clark *et al.*, [29]

This statement represents a vision for the innovation of technologies at the edge. However, this is a forward looking goal aimed at providing recommendations for future research. The thesis of this work supports this vision, focussing on mechanisms to better enable community-area networks at the edge.

However, there is a tension between the increasing demands of users and the capability of existing network technologies that provide access to Internet connectivity. Users want to maximise the value for money that they receive from any product or service that they purchase. This leads to an increased demand to push the existing capabilities of connectivity and associated hardware to their maximum.

The evolving use of the copper local-loop infrastructure is a good example of this tension. Exploiting the local-loop to avoid the costs of laying new data network cabling has driven the evolution of analogue modem technologies and then lead to the development of domestic ISDN and xDSL services specifically designed for digital connectivity. With the ever increasing need and desire for faster connectivity from users, xDSL services now offer multi-megabit data rates using the same physical infrastructure that at one time only offered a few 10s of Kb/s. Research continues into pushing further the limits of the existing local-loop. However, the legacy of the installed-base — features such as poor cabling installations, distance from exchanges and sometimes the slow enabling of exchange equipment — still restrict service provisioning for some users. Additionally, although the data rates offered by such wide-area connectivity are approaching the lower end of those possible in the local-area (a few Mb/s), they still fall well short of those possible at the higher end (up to 100Mb/s with proprietary extensions to 802.11g, and up to 1Gb/s with Ethernet).[1]

The wide-area wireless data market is less mature than the wide-area wired market thus, disparities between local-area and wide-area data rates are more pronounced. Wireless wide-area connectivity currently offers data rates of a few 10s of Kb/s with plans for 3G systems to offer a few 100s of Kb/s (or perhaps a few Mb/s at best). Yet many advanced services readily could be made available if these data rates were to improve. Therefore, even as connectivity technology advances in the mobile device arena, services also shall advance and so users will have increasing an expectation and desire higher data rates.

Community-area networking initiatives have begun to emerge as a result of a number of factors: the growing disparity between data rates available in the wide-area and in the local-area; the increasing demands and expectations of users; and the increasing availability and the decreasing cost of network-capable consumer equipment. As the number of such initiatives has grown, affiliations have been formed to promote their use and growth [11].

1.3 Thesis of Research

The thesis of this work is that ***users should collaborate to exploit fully their available (inter)network connectivity; by doing so they benefit from a higher throughput of data and a greater degree of robustness in connectivity within both fixed and mobile environments.***

1.3.1 Problem Domain

The problem domain consists of two main types of environment: a local community, within a geographically localised neighbourhood; and an ad hoc group of mobile users, within proximity of each other.

In both environments, users own one or more devices, networked together into either local-area or personal-area networks (using wired or wireless 802-standards technology) that reside at the edge of the Internet. However, local community users have relatively higher data rate Internet connectivity (e.g. using wired DSL or cable modems) than mobile users (e.g. GSM, GPRS and 3G).

As the number of collaboration efforts between the members of a community begin to increase and to intersect, we refer to the creation of a *coalition* within the community and the formation of a *Coalition Peering Domain (CPD)*.

1.3.2 Main Suppositions

Structure must be added to otherwise ad hoc collaboration agreements between peering edge inter-networks. Despite its initially disruptive effect on the status quo, this structure, the Coalition Peering Domain, provides direct benefits to the capacity of wide-area connectivity available for all collaborating members.

The premise of this work may be broken down into three main suppositions:

Hypothesis I The required structure enables transparency of communication between the members of the Coalition Peering Domain and may be provided through co-ordinated addressing and routing mechanisms. By enabling such transparency of communication, Coalition Peering Domain members experience the benefits of improved wide-area connectivity capacity.

Hypothesis II Within dynamically changing mobile environments, collaborating members may arrive and leave, or may gain and lose wide-area connectivity frequently. The structure provided by the formation of a Coalition Peering Domain provides robustness, allowing collaborating members to maintain better connectivity to each other, and to the wide-area.

[1] However, some countries including South Korea and Japan, have more advanced wide-area DSL connectivity that provides speeds of 100Mb/s and higher to the home.

Hypothesis III The structuring of connectivity agreements in a distributed manner forms an open system; one that can be applied to a diverse set of scenarios in which there are multiple levels of heterogeneity in the wide-area connectivity that is available to members.

These suppositions break down the main thesis of this work and test each individually.

1.3.3 Testing the Hypotheses

The first hypothesis involves an architectural design and reference implementation exercise to validate the main premise. It provides a proof–of–concept demonstration to show the relationship between wide-area data rate and membership within a Coalition Peering Domain, and thus the possible higher data throughput benefits of CPD membership for members. The measure of success for this hypothesis shall be the observation of a directly proportional relationship between wide-area aggregate data rate and membership within a Coalition Peering Domain. This is in comparison to the wide-area data rates that can be observed in the absence of a CPD.

Validation of the second hypothesis extends on the first by showing the benefits of Coalition Peering Domain membership, over longer term operation. It provides an analysis of the relationships between the size of a Coalition Peering Domain, the distribution of membership type within it (the proportion of edge and internal members), and the wide-area reachability by members as the distribution of membership type changes. This demonstrates that there is robustness in the wide-area connectivity as the distribution of membership type changes within a CPD over time. The measure of success for this hypothesis shall be the observation of a directly proportional relationship between wide-area reachability and size of CPD edge-membership.

The third hypothesis provides a validation of the architecture design within a wider context. This involves an exercise to deploy an implementation of the architecture within an environment where there is heterogeneity in the wide-area connectivity available to members. This demonstrates the applicability of a CPD within a wider context. The measure of success for this hypothesis shall be the demonstration to show that the previously observed relationships are reflected within the wider-context deployment.

The work-plan provided in Appendix C and detailed experimental outline provided in Appendix D break these down further, each to be evaluated thoroughly in an individual chapter within the final dissertation of this work.

1.4 Research Contribution

The contribution of this work is the introduction of structure and control into the problem domain. The work proposes a new architectural element, the ***Coalition Peering Domain (CPD)***, in which collaborating users together form a larger inter-network, residing at the edge of the Internet. This allows them to utilise more effectively and more efficiently their local connectivity resources.

A further contribution is the broader demonstration of how a CPD may benefit applications and scenarios beyond the problem domain.

2 Mechanisms Available Currently for Community Networking

Although existing mechanisms for addressing and routing are employed by individuals to administer existing community networking initiatives, such mechanisms do not provide an optimal solution. This is because community-area networking initiatives exhibit some key characteristics that distinguish them from existing classes of networks.

2.1 Administrative Responsibility

The community networks that result from existing initiatives, are essentially multi-homed, with many ingress and egress connections to the wide-area. Administrative responsibility is *distributed* across the community. Thus, they do not represent a single Administrative Domain (AD) that is under the control of a single organisation or entity, but rather a *collaborative* group of such entities. However, existing mechanisms for addressing and routing are designed to operate in network environments where administrative responsibility is not distributed. Formalising the relationships between community members is therefore not feasible using existing techniques alone.

2.2 Deployment and Maintenance Complexity

Proposing to employ existing BGP mechanisms [43] between community members is an extremely complex and heavy-weight solution. It requires first that each community member be allocated an Autonomous System (AS) number. This poses great problems for the existing Internet infrastructure, because, even the emergence of a small number of such communities would lead to a big explosion in the lengths of autonomous system paths and the numbers of routing table entries. Such back-pressure onto the core network is made even worse by the transient nature of community members' equipment: community members may switch on and off arbitrarily their equipment. Magnified by the explosion in the number of ASs and route entries, the resulting flapping of routes would lead to severe stability problems throughout the infrastructure.

Secondly, the configuration and maintenance of BGP requires a level of knowledge and expertise that is very unlikely to be available to most community members. Misconfiguration is already a source of problems within the existing Internet infrastructure [38], leading to unnecessary routing load and instability in the core routing tables. Poorly configured systems within community-area contexts would serve only to add to load, and propagate further any stability problems throughout the network.

Finally, employing existing BGP mechanisms requires that the equipment used is capable of supporting the relevant protocol and policy systems; this is particularly infeasible within resource-poor mobile or personal-area network environments, and would effectively exclude them from participating in such community network initiatives.

2.3 Multi-Homing

Traditional intra-domain ad hoc routing mechanisms [12] have focused on finding the single most efficient route on a source–to–destination basis, where the destination may be either inside or outside the local domain or ad hoc network. This models the domain or ad hoc network as a single AD that is, either disconnected, or a direct extensions of a larger infrastructure. This in turn requires them either to discover efficient routes to a very wide set of destinations, or to route towards a specific designated domain or network gateway (representing a single point of failure).

However, the multi-homed nature of the newly emerging class of community networks allows them to be seen as composite, multi-homed, virtual entities. They reside at the edge, but remain connected to, the Internet.

2.4 Connectivity Sharing

In the most basic form, connection sharing involves the simple routing of packets from multiple incoming interfaces to one outgoing interface; in more complex cases, it involves the use of multiplexing mechanisms to be applied to multiple incoming flows.

The increasing diversity, capability and affordability of network-capable devices have led to users owning multiple devices that are networked together. This has in turn led to an increasing demand for the capability to access the Internet from any of these devices at anytime and from anywhere. In most cases however, users have only single subscription-based connectivity to the wide-area Internet, available via only one or two of these devices.

Consumer operating system platforms (such as Windows and MacOS X) enable very simple local network set-up in a plug-and-play manner and configure 'connection sharing' automatically through a combination of Network Address Translation (NAT) and automatic address allocation using the Dynamic Host Configuration Protocol (DHCP).

The Linux-based netfilter and iptables tool-set [21] provides a more manual mechanism for configuring connection sharing using ‘masquerading’ and ‘port forwarding’. System administrators must define and activate the appropriate iptables rule-sets.

Hotspots share the characteristic of multiplexing traffic from many devices onto a single back-haul Internet link. However, they operate on a wider, commercial scale that involves many users and a larger number of devices.

Users may also choose to share their connectivity with others while not using it themselves. Such idle-time use aims to better utilise a connection by sharing it when it is temporarily idle, by treating the local device as a temporary gateway. An example is the 7DS Peer-to-Peer Information Dissemination and Resource Sharing system [40] that provides a mechanism for self-organised connection sharing. Load balancing mechanisms are also provided in 7DS, based on the selection of single (least loaded) gateways rather than distribution across multiple gateways.

So, although users may have potential access to multiple distinct wide-area connections, they are unable to easily utilise them all. This is because existing mechanisms for connectivity sharing tend to concentrate on sharing a single connection between multiple machines and do not take into account the possibility of utilising multiple connections simultaneously.

One variation of the existing connectivity sharing model is the ‘MAR commuter mobile access router’ [44], which provides an architecture for aggregating multiple heterogeneous types of wide-area connectivity. However it also focusses on a *hotspot* model (albeit multi-homed) of access with the placement of a ‘MAR’ device in moving vehicles. The device provides a range of local connectivity access (wired and wireless) for commuters. It is connected to the wide-area via multiple wireless interfaces, which it uses “simultaneously, to build a better combined wireless communication channel” and to provide bandwidth aggregation; it appears as a NAT box. However, this relies on all local users gaining wide-area access via a single provider (i.e. the MAR device) and thus represents a single point of failure. Again, it does not take into account the possibility that individuals may have some wide-area connectivity that could be better utilised.

2.5 Establishment of Trust

In their discussion on changes in the Internet since its inception, Clark *et al.*, [30] state that “... users don’t trust each other. The users of the Internet no longer represent a single community with common motivation and shared trust.” However, in the past, ad hoc and opportunistic networking approaches have focused on the automated discovery, negotiation and routing between neighbouring nodes that are all assumed to trust each other. This cannot be applied to the community network environment because it is very unlikely that members would be willing to trust all others unconditionally.

On the other hand, the newly emerging class of community-area networks does reintroduce the notion of community relationships on a local scale and within them, members *must* trust each other to some degree, for without this, such community-area networks cannot be formed. The newly emerging class community networks are therefore organised instead at the human level. This may be through either personal meetings or other forms of out-of-band interaction. This implies a basic, local, level of trust before any peering agreements can be reached, so, a level of trust is implied. However, this type of co-ordination is unsustainable in the long term and inhibits the future evolution of such community networks. These out-of-band interactions limit the efficiency with which community networks can be formed and may expand, requiring human-level intervention at every stage.

2.6 Discovery

Many of the existing community-area networking initiatives [7, 3, 4, 5, 10, 15, 17, 18] rely on growth of the network through mailing lists, member meetings, forums and events.

One such initiative, the Consume project [7], is “a collaborative strategy for the self provision of a broadband telecommunications infrastructure”. The project provides a web-based node database: a point of reference for prospective node owners to discover existing nodes within their local area. To be included in the node database, node owners must first register their node with the node database, providing details of its geographic location and its operational status (e.g. ‘speculative’, ‘testing’, ‘operational’ or ‘disabled’). Once registered, node owners must set-up their equipment and contact nearby nodes to establish peering agreements. Nodes may be connected together using any networking means available to the owners (e.g. either wired or wireless), and node owners may provide as many services as they wish. NodeDB [15] is another initiative that provides similar mechanisms for locating members.

Unfortunately this is insufficient to promote the growth of such community networking and leaves such initiatives well below any critical mass. They remain a fringe activity undertaken by experts and enthusiasts only.

The CUWiN project [6] has developed and released open source software to enable wireless community-area networking. The aim of the project is to allow "users to buy bandwidth in bulk and benefit from the cost savings". This project moves substantially towards easing the effort to establish oneself as a member of a community network. The project offers an installation CD that simplifies and automates all technical set-up and configuration, including "loading the networking operating system and software, sending out beacons to nearby nodes, negotiating network connectivity, and assimilating into the network". However, this level of automation comes at a price: the lack of local control available to individual users. This is unlikely to be acceptable to all individuals wishing to participate in the newly emerging class of community-area network initiatives.

The CUWiN approach relates closely to existing ad hoc networking approaches where connectivity between nodes is open and fully-automated, through the transmission of beacons; new nodes transmitting beacons are incorporated automatically into the mesh network. Thus the resulting mesh architecture appears to become a single edge network that extends the larger Internet, providing similar single-path source–to–destination routing mechanisms for traffic. Additionally, the CUWiN software appears to require higher-powered machines rather than potentially low-powered mobile devices. This limits the range of possible deployment scenarios, excluding any involving lower powered and mobile devices.

2.7 Self-Provided Networks

The aforementioned projects are just some of the many community networking initiatives in existence. FreeNetworks.org [11] "is a voluntary cooperative association dedicated to education, collaboration, and advocacy for the creation of free digital network infrastructures". It is an "affiliation between Community Wireless Networking Projects around the globe" and provides a portal for news about, information on, and references to free networking initiatives that have been formed around the world.

FreeNetworks.org provides a peering agreement document that is based on the Pico Peering Agreement v1.0 [19]; it defines a FreeNetwork as "any computer network that allows free local transit, following the guidelines of [their] peering agreement". The peering agreement provides a common method for formalising the agreements made between peers. This is a very valuable asset that may help support community-area networking initiatives because it provides a common base from which peering can occur, without placing substantial restrictions or dictating to individuals what they must or must not provide.

2.8 Summary of Deficiencies in Existing Community Networking Mechanisms

Not only do many community-area networking initiatives exist already, but the growing affordability of wireless-capable networking equipment provides encouragement for such initiatives meaning they are likely to grow.

The characteristics of this newly emerging class of community-area networks: distributed administrative responsibility; a requirement for ease of deployment and maintenance; their multi-homed nature providing higher potential aggregate bandwidth; distinguish them from existing classes of networks. Existing networking mechanisms do not provide optimal solutions to enable easily such community-area networks.

The growing number of community-area networking initiatives reinforces the popularity, and go some way toward enabling this newly emerging class of network. However such activities for the most part still rely on traditional networking mechanisms. Therefore they remain very much as fringe activities undertaken only by experts and enthusiasts.

2.9 Towards Better Mechanisms to Enable Community Networking

The thesis of this work aims to investigate, develop, and deploy a more suitable set of mechanisms that better enable this newly emerging class of network. To test the main hypotheses outlined in Section 1.3, a proposed dissertation outline and work-plan have been formulated and are provided in Appendices B and C respectively. To reach these goals, a number of experiments have been outlined in Appendix D.

3 Risk Assessment

There are a number of possible risks that may affect the completion of this dissertation. This section outlines these risks, evaluates the severity of the risks, and lists actions to mitigate them.

3.1 Availability of Equipment for Experimentation

The set of experiments required to validate the thesis of this work involve the set-up of a number of nodes with varying topologies and with varying degrees of complexity. The majority of these experiments are to be undertaken using the UCL Department of Computer Science Heterogeneous Experimental Network (HEN) [39, 36].

Possible delays in obtaining sufficient equipment to build-up the HEN testbed may cause work-plan timescales to slip significantly.

However, this risk is somewhat mitigated because a number of alternative items of equipment are available for use during experimentation. Three Linksys WRT54g wireless routers [13] are available for re-programming as necessary. Additionally, a further seven new high-performance compute nodes shall be available for use before the end of 2005 for further experimentation and for use to perform simulations.

The risk level may therefore be classed as low.

3.2 Similar Work

The possibility of identical work being undertaken and published independently is a risk that must be taken into account during any research project.

Community-area networking is a relatively new, emerging area that has not yet seen significant published work. This is a factor that helps to strengthen the case for pursuing the thesis of this work. In addition to a publication in the London Communications Symposium (LCS 2005), we have received three sets of research paper reviews acknowledging that the ideas presented here provide a new and unexplored research area. Additionally, the technical aspects of this work have been endorsed by two further reviews during a research council funding proposal. This emphasises the novelty of this work.

As this is such a newly developing research area, there is significant scope, and there are many directions for any work being undertaken within the area. Therefore the risk of independent work within this area invalidating the work presented here is much lower than would be the case in a more mature research area. The work-plan provided in Appendix C schedules initial sets of experiments to provide results within a 4 months of submission of this report. This should further mitigate the risks.

The risk level may therefore be classed as medium.

3.3 Feasibility of Work-plan

The possibility of being unable to stick to the work-plan may hinder the progress of this dissertation. The main risk factor here is that the investigations for this dissertation are being undertaken on a part-time basis. Other pressures may from time–to–time affect progress.

However, potential intersections with existing projects within the department may help to mitigate this risk. For example, the IST-funded RUNES project [37] aims to enable systems with embedded sensors to communicate, opening up new areas of applications. The application of CPD mechanisms into such environments may help contribute to the goals of the project.

The risk level may therefore be classed as medium.

3.4 Quality of Design

The possibility that the architecture and protocol design does not provide a valid solution is a risk that must also be considered. However, due to the incremental nature of the architecture and protocol design methodology, this risk is significantly reduced. Any inconsistencies or shortfalls may be rectified incrementally.

The risk level may therefore be classed as low.

4 Summary

In summary, the contribution of this dissertation shall be a thorough investigation, design, implementation and deployment of an architectural element that enables groups or communities of individuals to better utilise their wide-area connectivity. The architectural element — the ***Coalition Peering Domain*** — provides structure to the the mechanisms used by existing community-area networking initiatives.

Although the investigations begin by exploring the benefits within a neighborhood mesh context, there are a number of applications and scenarios beyond neighbourhood meshes that may benefit from the architecture being presented here. The investigations shall therefore validate first the architecture within the neighbourhood context, and then lead on to a validation of the architecture within a wider context.

A Challenges and Opportunities for Enabling Community Networking

This section highlights other work that has relevance to the thesis of this research.

A.1 Today's Internet: Service Provision versus Innovation

Connectivity to the Internet has very much become a commodity that many home users rely on for day–to–day needs, and many businesses rely on for daily operations. This somewhat overshadows the open, innovative, nature of the Internet infrastructure in its original form because there is a great pressure on providers to ensure the highest levels of reliability. As a result, there is no longer any room for the direct deployment of innovative, disruptive technologies within the Internet infrastructure. In his essay "It's About Connectivity Not The Internet!" [33], Frankston argues against this situation stating that the Internet "is supposed to be a medium for innovation" in a way that allows services to be connected at the edge without needing to worry about the middle. This argument is reaffirmed in the statement by Clark *et al.,* [29] quoted in Section 1.2.

Unfortunately, we find instead mechanism being placed frequently within the infrastructure (including firewalls, NATs and the blocking of ports). The reality however, is that much of the time, it is necessary. Although the Internet initially did consist of trusted peers, this is no longer the case and malicious behaviour from some users exploits any security flaws that were not foreseen when the Internet protocols were initially designed. This only serves to undermine the existing infrastructure. The priority now has become simply to sustain the infrastructure, patch any security holes and ensure that systems are not vulnerable to being compromised.

Fewer and fewer modifications to the existing infrastructure are made and any innovative, disruptive technologies that emerge are relegated to deployment within virtual networks and testbeds constructed atop the existing infrastructure. Unfortunately, such testbeds are not without problems. Once they reach any critical mass, they face exactly the same pressures for reliability.

To validate and demonstrate their visions for future communications technology, Clark *et al.,* [29] argue that "...it will be necessary to build some sort of prototype, testbed, or experimental infrastructure" and that this requires the networking community to reach some agreement on the sort of infrastructure needed for such experiments. However, in their report for the NSF Workshop on Overcoming Barriers to Disruptive Innovation in Networking [32], Peterson *et al.,* observe the tussles involved in both research-oriented testbeds and production-oriented testbeds. Although research testbeds can be very adventurous, their results do not provide any real indicative data because there is a lack of real traffic. On the other hand, although production testbeds carry real traffic, they must be conservative in their experimentation because of real user expectations on reliability. The same concerns were expressed also by some audience members following Huici's presentation on UCL's Heterogeneous Experimental Network (HEN) [36]. Strong encouragement was given to ensure that HEN is not allowed to fall under the type of administrative control that would undermine its usefulness as a flexible experimental testbed for real research.

Handley and Greenhalgh [35] observe that the Internet industry is not undertaking sufficiently the type of long-term research that would solve the increasing number of problems with the existing Internet architecture; this is partly because " researchers perceive insurmountable obstacles to experimentation and deployment of their ideas". They are not alone in expressing this concern. Schneider and Rodd's 'International Review of UK Research in Computer Science' [46] identifies under-provisioning of network infrastructure and systems research. The 'Computer Science and Telecommunications Board Committee on Research Horizons in Networks' report titled 'Looking Over the Fence at Networks: A Neighbor's View of Network Research' [47] comments on the need for additional systems-oriented research worldwide, which has been somewhat stagnated by the Internet infrastructure being unavailable for innovative (and possibly disruptive) research activities. These views are reaffirmed by Atkinson, Floyd, *et al.,* [26] who express concerns about the current state of funding for Internet research. They highlight that "Current funding levels for Internet research are not generally adequate, and several important research areas are significantly underfunded." They argue that the gap resulting from reduced levels of US Government funding since the mid-1990s has not been filled by the growing Internet industry as was expected. Instead, such commercial firms have tended to fund only the profit-maximising, low-risk areas that provide them "specific short-term economic advantage over [their] competitors."

This situation may, in part, stem from the networking community lacking a shared vision of the future [29]. As a result, point-solutions and work-arounds are applied to any problems discovered [32], or retro-fitted to existing protocols. Although this alleviates some of the problems in the short-term, the consequences of this strategy on the long-term well-being of the architecture are not good. They simply increase complexity and vulnerability to emerging threats.

A.2 Open Access Networks: Freedom versus Monopolistic Control

Battiti et al [27] argue that the high cost of services, the barriers for competition, and the inability of end users to be able to locally roam between providers, are a result of the vertical integration of networks. They propose that some of the problems can be solved through the provision of shared infrastructure.

The concept of warchalking [24, 25, 28] to advertise open wireless networks is an example of promoting the sharing (albeit unauthorised in most cases) of an open wireless local-area network, turning it into an access network.

More recent 'Open Access Networks (OAN)' involve physical access networks being shared by multiple operators. Examples include the StockholmOpen network [20] and the WILMA project [22], as well as the NoCat wireless community network [14] that distributes open source software to enable such networking initiatives.

The widespread growth of the Internet over existing telecommunications has meant that much of the end-user Internet access market has been controlled largely by the telecommunications companies that own the telephone network infrastructure. It can be argued that such monopolistic control has allowed prices to remain artificially high and the offered service range to remain prohibitively restrictive. As Battiti et al argue, it is important that the infrastructure is provided by a non-profit organisation because in most cases the OAN is monopolistic in nature and so "contrary to the openness concept". There is thus a danger that must be avoided if OANs are to be effective in avoiding the pitfalls that the Internet in its current state faces. This conflict is a difficult one to overcome; one that does not arise within a community-area network context. The distribution of administrative responsibility provides the necessary flexibility to allow individuals the freedom to establish alternative peering agreements and construct alternative infrastructures that meet their own specific demands.

A.3 User Friendliness: Customisation versus Out–of–the–Box Functionality

Consumers tend to spend a minimal amount of time modifying equipment and configuration if a specific need does not arise. For example, a study into the development of wireless networking in London [41] ran an 'Air Stumbling' (as opposed to 'war driving') experiment from a light aircraft with "... a directional antenna, a GPS and a laptop running network discovery program Netstumbler". It showed that out of 1525 nodes seen, 50% were 'open' and "... approximately 40% of access points are running with the manufacturers factory default SSID settings". While not a definitive measure, the figures seem to indicate that a significant portion of node owners may be non-technical and have found it sufficient to leave factory settings unchanged. This shows evidence of a potentially expanding market for out-of-the-box products aimed at allowing non-technical customers to participate in community-oriented networking activities, without needing to customise heavily their equipment, by offering auto-configuration and management systems to support CPDs.

By designing and manufacturing equipment that is flexible and simple to configure and to modify, manufacturers increase the likelihood of product success and benefit from the greater revenue that that success brings with it.[2]

This principle applies equally for software vendors. By designing and engineering software that specifically allows non-technical consumers to benefit easily from customised usage within a coalition network scenario, the software is likely to attract greater demand and produce greater revenue.

A.4 Transience: Intermittence versus Long-Lived Connectivity

Delay Tolerant Networking (DTN) [8] is a growing research area focussing on environments where end–to–end cannot be assumed. Emphasis is generally placed on whole "messages" rather than individual packets or message fragments with an objective to maximise the probability of message delivery in intermittently connected environments.

Assuming a degree of transience in the community-area (with users switching on and off their equipment), existing work in the DTN area may have some relevance to the work being investigated here. Two such examples are the HDNets and the HAGGLE systems.

The HDNet system [45] focusses on a highly dynamic multi-hop wireless network model in which clustering is used to allow higher powered 'mobile base stations' to forward data on behalf of lower powered 'mobile hosts'.

[2] An example is the Linksys WRT54g series wireless router (www.linksys.com/products/product.asp?prid=508&scid=35), which quickly became very popular on its release. Not only was it easy to re-flash the firmware on it, but the procedure remained an open option without any attempts from the manufacturer to prevent it.

The HAGGLE system [49] provides a "... networking architecture designed to enable communication in the presence of intermittent connectivity". It concentrates on sharing resources for data transmission on a store-and-forward basis rather than sustained aggregate connectivity.

Both systems focus on highly dynamic mobile environments wherein devices have neither reliable connectivity to the wide area nor sustained connectivity to each other with perhaps only brief opportunities (in the order of seconds or minutes) to forward data to each other. Additionally, HAGGLE is suitable only for delay-tolerant applications.

However, although nodes within a community-area networking context may to some extent be transient, characteristics such as strong focus on opportunistic and intermittent connectivity is not necessarily a relevant assumption in most cases. Such levels of intermittent connectivity as assumed by HDNet and HAGGLE are very unusual when considered in current situations. Community-area networking initiatives usually do involve long-lived connectivity between peers.

A.5 Traffic Distribution: Single-Path versus Multi-Path Routing

Existing mechanisms for routing, including those used by the aforementioned opportunistic networking systems, tend in general focus on single-path routing. However, the multi-homed nature of community-area network initiatives enables traffic to be distributed across the edge of the community-area network. This provides a mechanism for the load balancing of traffic at a packet or message fragment level.

The DIRAC software-based wireless router [50] provides a distributed router architecture composed of a Router Core (RC) and a Router Agent (RA). This may be useful inside a CPD boundary where routing functions can be shared and distributed, especially in scenarios involving inter-CPD communication. The merits of such an approach are to be further investigated.

Research into load balancing schemes within multi-homed networks dates back many years, including the work of Gibbens *et al.,* [34] into dynamic routing in multi-parented circuit-switched networks. The results of such research may have some influence on any protocol specification being undertaken within the community-area context. However, prior work in the packet switching domain has in general focused on single path source–to–destination routing, thus research involving multi-path routing shall be of particular importance.

A.6 Overlay Networks: Protocol Stack-Specific versus Technology-Specific

An overlay network can essentially be described as a network that is built on top of another network. Overlay networks may make use of protocols that reside either above, below or at the same layer of the given protocol stack being used by the underlying network. They are constructed using a mechanism known as *tunneling* which involves encapsulating the overlay network's packets within the underlying network's protocol or network infrastructure. The resulting overlay can therefore consist of neighbouring nodes, which when viewed from the underlying network infrastructure's perspective, are within separate networks and may be geographically dispersed.

Overlay networks are usually deployed for specific purposes; for example to provide a specific type of infrastructure that is built on top of another, or to provide a logically isolated infrastructure. In general, overlay networks possess the following set of qualities:

- They are built on top of an existing (underlying) network
- They make use of tunneling through the underlying network for communication at the same layer between nodes within the network
- Neighbouring nodes within these networks are not necessarily neighbouring nodes within the underlying network thus, links between neighbouring nodes are virtual or logical, rather than direct or physical.

Overlay networks can be said to fall under one of two categories: "Protocol Stack-Specific" and "Technology-Specific". In general, protocol stack-specific overlays provide specific services to upper layers and are deployed to provide basic communications infrastructure. On the other hand, technology-specific overlays are in general born from a need either to provide a testing ground for new and emerging research technologies; or to provide a specific service (*on top of* existing infrastructure), due to a lack of capability or feasibility to deploy directly within existing infrastructure. For this reason, technology-specific overlays may be seen as *transitional* overlays that may later become integrated into existing infrastructure. Some examples of technology-specific overlays include:

6Bone Since general Internet routing equipment did not support the forwarding of IPv6 protocol packets, the IPv6 Backbone (6Bone) [1] was until recently, an overlay network deployed on top of the existing IPv4 Internet infrastructure (through IPv6-in-IPv4 tunneling) to test the IPv6 protocol implementations.

MBone Likewise the Multicast Backbone (MBone) [31] was an overlay network set up (using IP-in-IP tunneling) to test various multicast protocols and technologies. This was established because most Internet routing devices were unable to forward multicast data either through incapability or through device owner configuration policy.

ABone The Active Network Backbone (ABone) [2] provides a "virtual testbed for the active networks research program funded by DARPA ATO" [2].

A number of dynamic overlay and Virtual Private Network (VPN) configuration technologies are available, both commercial and open-source. These are based both on hardware and on software. Further details on some selected dynamic overlay network systems are provided in appendix E.

There currently exist no protocols that directly enable existing community-area networking initiatives. Therefore these existing systems tend to fall within the technology-specific overlay category. Such initiatives often involve the establishment of tunnels between peers, for providing specific services to each other (e.g., sharing resources).

B Proposed Dissertation Outline

C Work Plan

This section provides an outline schedule for completion of the dissertation.

C.1 "Chapter 2" — Background and Motivation

This task involves a wide analysis of the general research area and a thorough analysis into the motivations for community-area networks.

This task has been partially completed.

Estimated time to completion: 6 weeks

C.2 "Chapter 3" — State of the Art

This task involves a thorough analysis of existing systems relevant to community-area networking, an evaluation of their deficiencies and an identification of where the work of this thesis fits.

This task has been partially completed.

Estimated time to completion: 8 weeks

C.3 "Chapter 4" — The Coalition Peering Domain

C.3.1 Architecture Design and Protocol Specification

This task involves the core design of the overall Coalition Peering Domain architecture.

- Architecture and topology design and definition
- Protocol specification
- Protocol design validation

This task is in progress. Early specification descriptions have been included in Appendix F.

Estimated time to completion: 8 weeks

C.3.2 Analysis of Architectural Challenges

This task involves a thorough analysis of the architectural challenges faced for the design and deployment of the CPD. It also involves a thorough analysis of the impact of the CPD on the existing infrastructure.

This task is in progress.

Estimated time to completion: 4 weeks

C.3.3 Core Fixed-Environment Scenario Evaluation

This task involves a full validation of the architecture design and protocol specification showing evidence that users do benefit from collaboration with the formation of a CPD. This shall be undertaken through the implementation of the protocol and a combination of simulation and experimentation.

A detailed outline of the experiments to be undertaken for this task can be found in the Detailed Experiment Outline Section D.

This task will begin following completion of task C.3.1.

Estimated time to completion: 16-20 weeks

C.4 “Chapter 5” — Robustness within Mobile Environments

C.4.1 Impact to architecture and Protocol Design

This task involves a thorough analysis of the impact of disconnection to the CPD architecture. The aim is to show evidence that users are able to maintain a more robust level of connectivity to the wide-area through the multi-homed nature of the CPD. This shall be undertaken through a combination of simulation and experimentation.

A detailed outline of the experiments to be undertaken for this task can be found in the Detailed Experiment Outline Section D.

This task will begin following the completion of task C.3.3

Estimated time to completion: 28 weeks

C.5 “Chapter 6” — Applicability within a Wider Context

This task involves a validation of the CPD architecture within a wider context. The aim is to demonstrate that the CPD architecture can benefit a broad range of scenarios in which there is a heterogeneity in the available wide-area connectivity. This shall be undertaken through direct experimentation.

A detailed outline of the experiments to be undertaken for this task can be found in the Detailed Experiment Outline Section D.

This task will begin following the completion of task C.4.1

Extimated time to completion: 20 weeks

C.6 Dissertation Write-up

Completion of write-up and integration of final dissertation

Estimated time for completion: 20 weeks

D Detailed Experiment Outline

D.1 Metrics

The core metrics to be used for evaluation are:

m1 Average End-to-End Data Rate — Defined as: $\frac{Total\ Number\ of\ Bytes\ Transferred}{Time}$

m2 Average Throughput — Defined as: $\frac{NumPackets\ Received\ Remotely}{Total\ NumPackets\ Transmitted}$

m3 Average Packet Loss — Defined as: $1 - Average\ Throughput$

m4 Average End-to-End Latency — Defined as: $\frac{\sum_0^N umPacketsRTT}{NumPackets}$

m5 Average (Routing Table) Recovery Time — Defined as: the time taken for all CM routing tables to be updated with the correct routes following a change in membership.

D.2 Network Topologies

The experimental testbeds are to be built up incrementally, beginning with a basic two-node topology and expanding in size and complexity. These have been illustrated in Figure 1.

t1 This is the most basic topology involving only two peering nodes, both acting as CEFs distributing traffic between each other.

t2 This topology introduces a CIF into the basic set-up of t1. The two CEFs peer together and distribute traffic between each other. In addition to distributing its own traffic, one CEF peers with the CIF and distributes also the CIF's traffic to its neighbouring CEF.

t3 This topology builds on t1 by increasing the number of CMs that form a three-member CPD. All three CMs act as CEFs.

t4 This topology builds on t2 by increasing the number of CIFs. Two CEFs form the CPD-edge and distribute traffic arriving from multiple CIFs.

t5 This topology builds on t4 by increasing the number of CEFs to three.

Each individual topology includes a remote entity — a node residing outside the CPD with which communication is to be undertaken. For each topology, a control set-up shall be establisged where coalition-based peering is not used, thus traffic is not distributed between neighbouring peers. This shall provide a benchmark against which the CPD architecture may be evaluated.

Further advanced topologies shall be formulated on completion of the initial experimentation phase and result analyses.

D.3 Test Applications

The set of application tests to be undertaken range from basic connectivity establishment tests to more advanced multi-party conferencing.

a1 Traceroute: for basic reachability testing

a2 File Transfer — Single-User: A single file transfer process from a CM to the remote entity.

a3 File transfer — Multiple-User: Multiple file transfer processes from each CM to the remote entity.

a4 Video transmission — Uni-directional, Unicast, Single CM: from a single CM to the remote entity.

a5 Video conference — Bi-directional, Unicast, Single CM: between a single CM and the remote entity.

a6 Video transmission — Uni-directional, Unicast, Multiple CMs: from multiple CMs to the remote entity.

a7 Video conference — Bi-directional, Unicast, Multiple CMs: between multiple CMs and the remote entity.

Further advanced application tests involving multicast transmission may be formulated following the completion of the initial experimentation phase and result analyses.

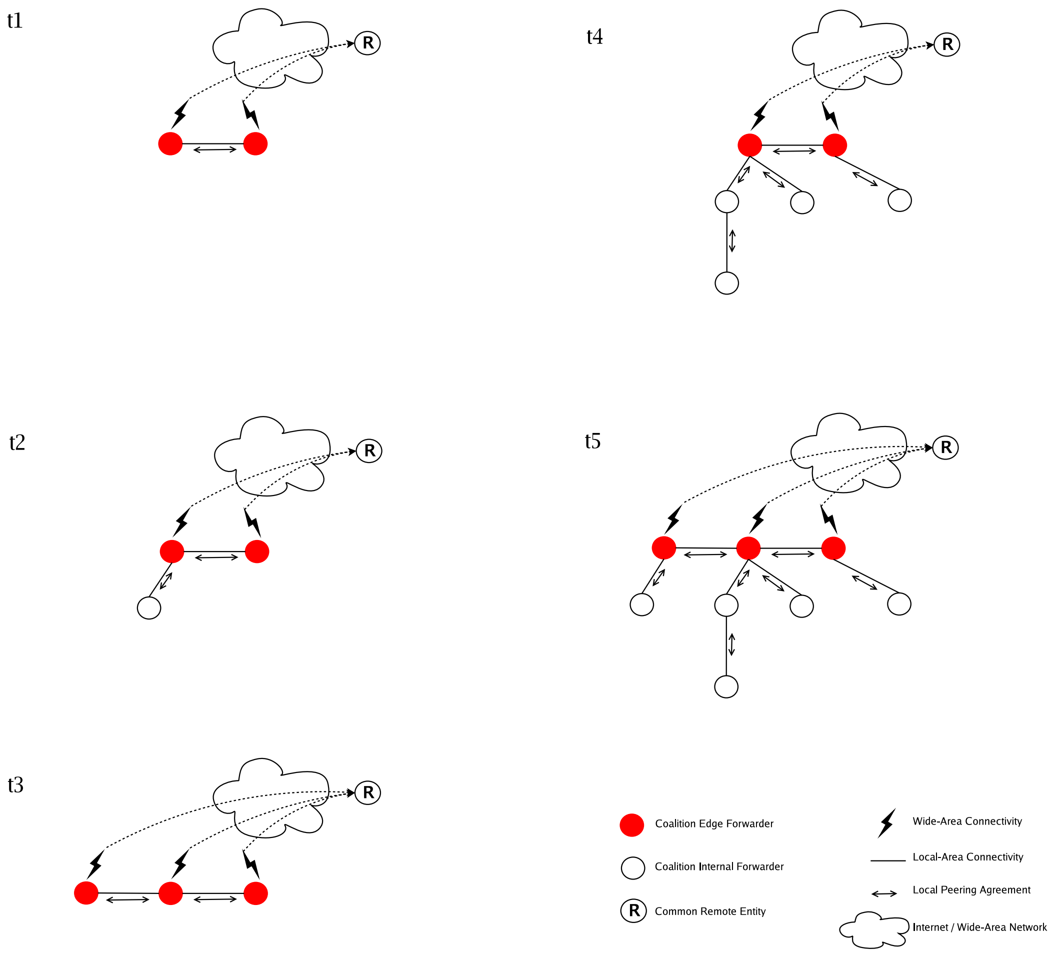

Figure 1: Experimental Network Topologies

D.4 Core Fixed-Environment Scenario Evaluation

D.4.1 Outline

This involves the combination of each application test run sequentially across each of the planned network topologies.

D.4.2 Expected Results

The expected results are that for each combination, metrics m1 and m2 are higher, and metrics m3 and m4 are lower when compared against control topology conditions.

D.4.3 Evaluation Criteria

The criteria used for evaluation are:

c1 A comparison for each of the metrics against the number of CEFs involved while only one CEF generates traffic for distribution.

c2 A comparison for each of the metrics against the number of CEFs involved while multiple CEFs generate traffic for distribution.

c3 A comparison for each of the metrics against CEFs represented as a ratio dependant on the size of the CIF-chain below them.

D.5 Impact to architecture and Protocol Design

D.5.1 Outline

This involves the combination of each application test run sequentially across each of the planned network topologies with the added actions of CEF disconnection and re-connection mid-way through each run. This shall initiate a change in the running topology causing re-routing and re-negotiation to occur and so provide an evaluation of robustness and adaptability. These actions simulate environments where disconnections and re-connections are frequent.

D.5.2 Expected Results

The expected results are that for each combination, metrics m1 and m2 are higher, and metrics m3, m4 and m5 are lower when compared against control topology conditions.

D.5.3 Evaluation Criteria

The criteria for evaluation remain identical to those used during the Core Fixed-Environment Scenario Evaluation D.4.

D.6 Applicability within a Wider Context

D.6.1 Outline

This involves the instantiation and deployment of the CPD protocol mechanisms within a non-neighbourhood context. The deployment scenario chosen is the resource-constrained, hybrid (mobile and fixed) environment that would be available either to authorities arriving at the scene of an emergency, or to relief workers arriving at the site of a natural disaster. The deployment shall therefore incorporate a network of heterogeneous devices with varying connectivity to the wide-area.

D.6.2 Expected Results

The expected results are that metrics m1 and m2 are higher, and metrics m3, m4 and m5 are lower as observed by each member device within the scenario compared against the control scenario where no CPD exists.

D.6.3 Evaluation Criteria

The base criteria for evaluation remain identical to those used during the Core Fixed-Environment Scenario Evaluation D.4. In addition to this, further variations shall be applied to the set of topologies used to emulate the scenario more realistically.

E Dynamic Overlay and VPN Systems

This section highlights some selected overlay and VPN systems that have relevance to community-area networking.

E.1 Hardware-Based Solutions

A number of manufactures have product lines that include 'hardware VPN solutions'. For example, NSGDatacom's product line includes the ADI Assured Digital range [16]. Such solutions are essentially devices such as switches, gateways, or routers, with multiple network interfaces. They contain dedicated hardware and firmware capable of automating the establishment of IPSec VPN tunnels with remote sites. Additionally, capabilities beyond basic VPN tunnel establishment are available, such as authentication and authorisation between devices as well as the configuration of security policies.

In general, the aim is to provide a device, not unlike a NAT or firewall device, that simply can be connected to the the edge of a (local-area) network, such that it acts as a gateway for connecting securely to remote networks. As such, these devices provide a level of dynamism in the configuration and establishment of IPSec VPN tunnels that has not until recently been available both at the hardware and the software level. Although such hardware products often provide proprietary overlay/VPN solutions that more suited to commercial markets, such features may also benefit the community-area, as exemplified by the emerging devices for home users that do include these features.

E.2 Software-Based Solutions

The traditional software configuration of an IPSec VPN tunnel involves a manual process whereby administrators at each site must configure their local VPN tunnel end point with the necessary software, addressing, encryption keys and firewall policies, before the VPN tunnel can be activated.

Various software solutions now exist that try to automate this process through the provision of software daemons and graphical interfaces for simpler tunnel definition. They share a few basic elements:

- Tunnel management interface
- Encryption key distribution
- Address allocation and configuration distribution
- Software-initiated or script-initiated tunnel deployment

E.2.1 USC/ISI X-Bone

The X-Bone is "a system for the dynamic deployment and management of Internet overlay networks" [48]. It has been developed at the University of Southern California (USC) Information Sciences Institute (ISI). "The X-Bone discovers, configures, and monitors network resources to create overlays over existing IP networks" [23]. Its goal is to reduce configuration effort and increase network component sharing.

The X-Bone extends current overlay management by introducing dynamic resource discovery, monitoring and component reuse. Nodes (hosts or routers) can participate simultaneously in multiple overlays, not only at the same level, but also in a hierarchic manner. This provides the capability for supporting recursive or hierarchic overlays (overlays that are built on top of other overlays). It does not require any operating system-specific or application-specific modifications. It uses basic IP-in-IP encapsulation capabilities, existing implementations of dynamic routing, the Domain Name Service (DNS), and IPSec.

The X-Bone system essentially consists of two component elements:

Overlay Manager (OM) responsible for deploying and coordinating overlays

Resource Daemon (RD) responsible for coordinating the resources of individual network components

A web-based CGI-implemented user interface (the OM GUI) is provided for communicating with the OM to carry out the construction, management and and dismantling of overlays.

To establish an overlay, an OM sends out an overlay invitation. This is undertaken either in the form of an expanding multicast ring search, or unicasting to a pre-defined set of hosts provided to the OM GUI. RDs listen on a pre-defined port for invitation messages. On receipt of an invitation, an RD may respond to the invitation; this indicates its availability for inclusion into the overlay. Such a response from an RD is optional, so, RDs have some degree of control for deciding in which overlays they will participate.

The OM waits for a specified period of time and once the timeout has been reached, the OM proceeds with the overlay creation process. This involves first checking that the correct number of RDs have responded (otherwise the overlay creation fails). If it has received more responses to the invitation than required, the OM selects arbitrarily the required number of nodes. To establish the overlay, the OM determines first the tunnel endpoint addressing and routing table entries that must be configured at each RD. These determinations are based on a number of parameters obtained from the OM GUI. These parameters include details of overlay topology: star, ring, linear or user-specified (using a special definition language); and details of encryption and authentication algorithms for tunnelling. The OM then sets up X.509 encrypted TCP/SSL connections to each of the selected RDs so as to transmit the relevant configuration information. On receiving the relevant configuration information, the respective RDs activate their tunnel interfaces.

The creation of overlays within the X-Bone is carried out using a *two-layer tunnelling* mechanism for each level of overlay. The first layer provides a virtual link layer on top of which the network layer overlay is built. This enables the use of multicast, dynamic routing, and IPSec, within the overlay because these intrinsically are network layer mechanisms. This method results in three IP headers in the case of an overlay constructed on top of the base network; the innermost header represents the end points within the overlay, the next header acts as a link layer, and the outermost header represents the tunnel endpoints within the base network.

E.2.2 DRDC DVC

The Dynamic VPN Controller (DVC) system [42] developed by NRNS Inc. for the Defence Research and Development Canada (DRDC) agency, provides mechanisms for dynamically negotiating, establishing, maintaining and dismantling IPSec-based VPNs, making use of secure, authenticated out-of-band channels. Its main feature is the entirely distributed nature of the infrastructure that is established, with each DVC site maintaining its own set of security and access policies to its local ‘participating network resources’. This is achieved by the establishment of ‘coalitions’ in which partners “need only maintain high-level information about each other — e.g. who they are and where they are on the network”. The establishment of coalitions between various peers means also that any given peer may be, at any time, a member of several different coalitions. This provides a highly secure and dynamic mechanism for VPN establishment with a great degree of control ensuring local security remains the priority.

The main concept and technology behind the system has been derived originally from the X-Bone [23] system for overlay deployment, but places subsequently a more detailed focus on policy. Like the X-Bone, the DVC is written largely in Perl.

Each coalition partner site runs a local DVC that is connected to a common wide area network. Coalition partners make their ‘participating network resources’ available through their respective local DVC. Each DVC maintains a local XML-based policy database, constructed through a Java-based policy editor tool, to dictate access to local resources. A web-based CGI-implemented user interface is provided for initiating and disabling coalition connections. To establish a connection when a partner makes a request to join a coalition, the local DVC initiates a connection to a remote DVC via SSL. The initiating DVC provides its security policies; these may be passed up to the DVC Operators at the remote sites. If the remote DVC Operator acknowledges the initiating DVC security policies, the remote DVC’s security policies are sent to the initiating DVC Operator. On acknowledging the remote DVC’s security policies, each DVC configures its local system to establish the required IPSec VPN tunnel. When a coalition partner’s access policies are modified within the local policy database, the DVC is notified and must re-negotiate the VPN connection terms with each of its current coalition partners. Similarly, if a coalition partner’s access is terminated, the remote DVC is notified and the relevant VPN tunnel is dismantled. In this way, a coalition connection is established between a pair of coalition partner sites wishing to communicate thus, a fully-meshed topology is formed.

The DVC succeeds in examining the use of policy enforcement in a very thorough manner. This helps to move policy management decisions to an autonomous level.

It provides dynamism in establishing and dismantling VPN connections through the use of the security policy negotiation system. However this comes with a cost: the level of static information required to know where other coalition partners are within the network.

F Early Draft of CPD Protocol Specification

F.1 Introduction

This document is an ongoing work in progress, defining the main protocol to enable the formation and operation of a Coalition Peering Domain (CPD).

F.2 Protocol Description

Figure 2 illustrates a number of collaborative efforts or ‘local peering agreements’ between pairs of community members. These peerings may be either as simple as links interconnecting different pairs of community members, or more complicated associations controlled through policy defined locally by the community members. As the numbers of such local peering agreements begin to increase and to intersect between community members, we refer to the creation of a *coalition* within the community and the formation of a *Coalition Peering Domain (CPD)*.

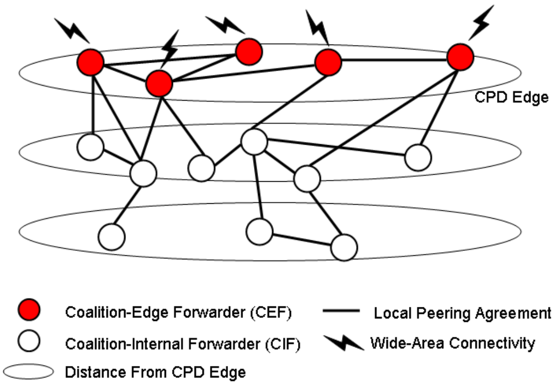

Figure 2: CPD Architecture

Each Coalition Member (CM) may represent an individual with either a single node, or a local network. Coalition members who have wide-area connectivity (or more generically, connectivity outside the CPD) form together the edge of the CPD and act as *Coalition-Edge Forwarders (CEFs)*; they are the CPD ingress–egress points, allocating some proportion of their external connectivity for this purpose. In the simplest case they may forward outgoing packets on their CPD-egress link. However, in a more interesting case they may forward some of these outgoing packets by ‘spraying’ (distributing) them, across the CPD edge, via their CPD-internal interfaces to other member CEFs within range, who then forward the packets outside the CPD. Thus outgoing traffic is distributed across multiple CEFs, so enabling a higher upstream data rate by aggregating multiple CM egress links. This type of wide-area connectivity aggregation is an example of collaboration between individuals for mutual benefit. This approach is useful when the local capacity between a number of CMs is greater than or equal to their individual egress capacity to a common remote entity.

Coalition members who do not have connectivity outside the CPD, or who choose not to make available their wide-area capability to other CMs, act as *Coalition-Internal Forwarders (CIFs)*. The forwarding of CPD-internal traffic (the traffic traversing between CMs) may be performed using modified forms of standard inter-domain or ad hoc routing protocols. In the example of collaboration for the purpose of wide-area connectivity aggregation, CIFs forward CPD-outbound traffic by directing it towards their ‘nearest’ CEF for CPD egress. This traffic can be sprayed across the CPD edge by the receiving CEF as described above, thus, CIFs may also benefit. Of course, CIFs may also use mechanisms for load balancing and take responsibility for spraying directly to multiple CEFs, depending on the physical connectivity of the CPD.

In this context, coalition members represent a reasonably static group of nodes or local networks that form peering agreements between each other. It is also possible for such connectivity to be extended to non-coalition members, for example mobile/roaming nodes travelling through the CPD. These may peer dynamically with a CIF or directly with a CEF as they pass within radio range.

The establishment of local peering agreements between CMs could be completely manual, but the intention is eventually to have some level of auto-configuration, based on secure authentication (e.g. PGP keys).

F.2.1 Negotiation

Simple Unicast

Each prospective CM wishing to form a local peering agreement, unicasts a CPD-PREQ message to a target. The CPD-PREQ recipient is given effective control of the negotiation process to decide:

- Whether to agree to the formation of the local peering agreement

- Whether to include the local peering agreement into one of its existing CPDs or form a new CPD

- The channel parameters of the local peering agreement

If the CPD-PREQ recipient does wish to proceed with local peering agreement formation, it unicasts, back to the requester, a positive CPD-PRESP message. On receipt of the positive CPD-RESP by the requester, the requester must send an initial CPD-ADV message on the required channel to confirm local peering agreement formation and thus local peering agreement formation is said to have succeeded. To maintain the local peering agreement, each peer-CM unicasts subsequently, at regular time intervals, a CPD-ADV message on the correct or negotiated channel.

If the CPD-PREQ recipient does not wish to proceed with local peering agreement formation, it unicasts, back to the requester, a negative CPD-PRESP message. On receipt of the negative CPD-RESP by the requester, local peering agreement formation is said to have failed.

The CPD-PREQ message contains:

- TBD

The CPD-PRESP message contains:

- TBD

The CPD-ADV message contains:

- E_L — the local egress capacity

- S_L — the local spray capacity

- $T_{L,R}$ — the trust level that local has for remote

- $CPD_{ID}, Port$ — the CPD ID and port to be used for communications

The responding peer has control of the peering process. It chooses whether to use a CPDid of a CPD it already belongs to (in which case the resulting local peering agreement becomes part of the existing CPD) or to generate a new CPDid (in which case a new CPD is created containing a stand-alone local peering agreement. In this way, CPD membership grows and evolves over time.

Problem: What happens if a central CEF exits the CPD leaving two parts of the CPD disconnected? They can't maintain the same CPDid can they? Initial thoughts are that they potentially could... so long as the Link State algorithm is able to propagate information quickly enough. In the future, if one of the CEFs/CIFs from one of the disconnected segments attempts to peer with a CEF/CIF from the other segment, and if receives a CPD-PRESP containing the CPDid to which

Multicast

All CMs and prospective CMs listen on a "well-known" CPD multicast channel $CH1$. Each prospective CM wishing to form a local peering agreement, multicasts a CPD-PREQ message on $CH1$. CPD-PREQ recipients are given effective control of phase 1 of the negotiation process to decide:

- Whether to agree to the formation of the local peering agreement

- Whether to include the local peering agreement into one of their existing CPDs or form a new CPD

- The channel parameters of the local peering agreement

If a CPD-PREQ recipient does wish to proceed with local peering agreement formation, it [uni—multi]casts, back to the requester, a positive CPD-PRESP message. On receipt of all the positive CPD-RESP messages received by the requester, the requester has effective control of phase 2 of the negotiation process to decide:

- Whether to agree to the formation of a local peering agreement with each positive responder

The requester must choose which (subset) of the responders it wishes to peer with, and must send an initial CPD-ADV message on the required channel to confirm local peering agreement formation and thus local peering agreement formation is said to have succeeded. To maintain the local peering agreement, each peer-CM unicasts subsequently, at regular time intervals, a CPD-ADV message on the correct or negotiated channel.

If a CPD-PREQ recipient does not wish to proceed with local peering agreement formation, it can choose to either ignore the CPD-PREQ or send, back to the requester, a negative CPD-PRESP message.

F.2.2 Addressing

Assumptions

- CEFs have egress IPv6 connectivity, thus a global IPv6 address allocated to them by their wide-area provider

- CEFs may possess an additional IPv6 /64 allocation

Addressing Functionality

CEFs use their own global IPv6 address (allocated to them by their wide-area provider) to identify themselves.

CEFs that do not possess a global /64 allocation cannot peer with CIFs. CEFs that do possess a global /64 allocation are said to have the capability to peer with CIFs. They sub-allocate one subnet to each CIF with which they peer.

CIFs may also peer with other CIFs. CIFs sub-divide their subnet (allocated from their upstream CIF or CEF and sub-allocate each of these sub-divided subnets to each downstream peer CIF. A CIF peering with multiple other CMs therefore is allocated multiple subnets: one from each of its upstream peers. They can choose the highest order allocation to deduce the shortest number of hops route to CPD-edge.

Problem: What happens when a CEF switches and becomes a CIF? — it would need to renegotiate with its peer CEFs and be allocated a subnet from its upstream CEF's /64 allocation.

Alternative NAT-based Addressing Functionality

Without the assumptions of IPv6 and /64 allocations, it is possible for CEFs to operate a NAT for address allocation to CIFs that peer with them. They can then sub-divide their NAT address pool across their CIFs. CIFs can thus sub-divide and sub-allocate their allocation to downstream CIFs in the same way as described above for IPv6 /64 sub-allocation.

F.2.3 Routing

CEFs operate a Link State algorithm between themselves (i.e. across the CPD-edge) to ensure that each CEF has a map of the CPD-edge topology and knows how to reach the different CEFs spread across the CPD-edge. (REFER TO Lakshmi's secure routeing talk and paper)

With the /64 allocation approach:

The Link State algorithm will include information on all CEFs. The Link State algorithm will also include information on any CIFs that have their own /64 allocations in use downstream. This is the case only when a CEF switches into a CIF. When this happens, it is allocated a subnet by each of its peer CEFs, but it also continues to send them the relevant link state information for its downstream CIFs therefore the routeing to that /64 allocation can continue seamlessly through it.

With the NAT approach:

There is no need for the Link State algorithm to include any information about the CIFs. This is because they reside within a CEF NAT so traffic reaches the CEF first and then the CEF realises that it should be routed to one of its private networks.

Problem: If a CEF becomes a CIF, it must switch over to its newly allocated NATed addresses from its peer-CEFs and sub-allocate these to its downstream CIFs before the downstream CIFs can continue.

CPD-ingress traffic routeing

TBD

CPD-egress traffic routeing

TBD

F.2.4 Trust

A simple distributed trust mechanism can be used across the CPD to self-regulate CM behaviour. This section outlines the basics that such a mechanism would involve. This work has contributed to the development of a more complete, generic and independant model by Daniele Quercia, and is due to be presented during the ACM Symposium on Applied Computing (SAC) in April 2006.

Assumptions

- For UDP packet forwarding, we solve nothing — it is no different to current udp traffic on the Internet

Trust Mechanism Notation

- P_x = A peer CM ‘x’

- $T_{x,i}$ = The level of trust that P_x has in P_i

- $ADV_{x,i}$ = A CPD-ADV sent by P_x to P_i

- $LPA_{x,i}$ = A local peering agreement existing between P_x and P_i

- $C_{i,x}$ = A credential containing the trust that P_i has in P_x.

A credential is composed of: $(T_{i,x}, Timestamp, \{Hash(T_{i,x}, Timestamp)\}_{SKi})$ - C = The total number of credintials

- $*$ = multiplication

Trust Propagation Mechanism

Trust Bootstraping: All peer CMs P_i save the set of best credentials $C_{j,i}$ that they receive from peer CMs. (These are received through CPD-ADV messages).

During local peering agreement formation between P_x and P_y, P_x sends to P_y the level of trust it has in P_y ($T_{x,y}$). P_x also sends to P_y the set of best credentials $C_{i,x}$ that it has collected from the set of nodes P_i.

P_y extracts the credentials and computes its own initial trust level towards P_x. This is:

- $\propto w_{y,i} * T_{i,x}$

- $\propto C$

- $\propto The freshness of all credentials$

Trust Updating A CM P_x may update the level of trust it has in its peer CM P_y under two circumstances:

1) P_x is not receiving all its ACKs relating to traffic it has sent

2) P_x receives a degraded trust notification $T_{w,x}$ from any of its other peer CMs P_w

There are two possible approaches: one where state is kept (aligned with TCP window), and another without state (proportionally punishing all and propagating across CPD).

F.2.5 Providers’ Side Services

It is ultimately the providers who allocate global IPv6 addresses to their subscribers and configure the appropriate routing to the IPv6 /64 allocations of their subscribers.

Providers may also make available a “Reverse Aggregator”: a server that is provides reverse packet spraying to a specific CPD. Such a server may be provided as a ‘well known’ resource in the same way that ISPs provide their subscribers, to configure their local connectivity, DNS server and default gateway addresses.

Reverse Aggregator Functionality

If a Reverse Aggregator is available, then the preferred method for a CM is to establish an IP-in-IP tunnel to the Reverse Aggregator, and to route all its CPD-egress traffic through the tunnel. The Reverse Aggregator is responsible for forwarding this traffic on. As the Reverse Aggregator is in the reverse path of some packets destined for CPD-ingress, it is responsible for reverse spraying packets across the CPD edge. The Reverse Aggregator carries out load balancing by holding some state on the CPD-egress traffic and reverse spraying proportionally. So it would increase a CM counter for each CPD-egress packet from that CM, and decrease the CM counter for each CPD-ingress packet it sprays into the CPD through that CM.

- $CM_i \rightsquigarrow 0 : +1 perCPD - egress, -1 perCPD - ingress$

Tunnel Parameters are:

- IPv6 end-point address

- Key == CPD_{ID}

Therefore, the Reverse Aggregator can distinguish groups of CMs/tunnels that belong to a single CPD, and reverse spray accordingly.

If a Reverse Aggregator is not available, then the default method is for a CM to transmit all its CPD-egress traffic on its normal wide-area link. As this node is directly in the reverse path of some packets destined for CPD-ingress, it will bear a greater burden of ingress traffic because no ingress spraying is in operation. However, this may be offset by the assymetry of its wide-area connection.

Therefore, this mechanism does not require that all CMs within a CPD must be subscribed to a single provider. A Reverse Aggregator may be made available by a provider and reverse spraying may be performed to the *subset* of CMs tunnelling with the same CPD_{ID}. CMs who are not subscribers of that particular provider may make use of their own provider's Reverse Aggregator, or may make use of their normal wide-area link.

Problem: When a CPD is segmented.... packets may be reverse sprayed across the different segments and end up in the wrong CPD segment. See addressing problem above relating to CPD segmentation.

F.3 Summary

TBD

Acknowledgements

Daniele Quercia — Critique for initial trust mechanism. Continued development, refinement and work on a distributed trust mechanism applicable to CPD scenarios.

References

[1] 6BONE testbed for deployment of IPv6. http://6bone.net/.

[2] Active Network Backbone (ABone). http://www.isi.edu/abone/.

[3] Austin Wireless. http://www.austinwireless.net/.

[4] Bay Area Wireless Users Group. http://www.bawug.org/.

[5] BOUNDLESS Community Wireless Network. http://boundless.coop/.

[6] Champaign-Urbana Community Wireless Network. http://www.cuwireless.net/.

[7] Consume “trip the loop, make your switch, consume the net”. http://consume.net/.

[8] Delay Tolerant Networking Research Group. http://www.dtnrg.org/.

[9] Elektrosmog. http://www.elektrosmog.nu/.

[10] free2air.org. http://www.free2air.org/.

[11] FreeNetworks.org volunteer co-operative association. http://scoop.freenetworks.org/.

[12] IRTF RRG Ad hoc Network Systems Research Subgroup. http://www.flarion.com/ans-research/.

[13] Linksys WRT54G Wireless-G Broadband Router. http://www.linksys.com/products/product.asp?prid=508&scid=35.

[14] NoCat. http://nocat.net/.

[15] NodeDB. http://www.nodedb.com/.

[16] NSGDatacom ADI Assured Digital Product Range. http://www.nsgdata.com/adi/index.html.

[17] NYCwireless. http://nycwireless.net/.

[18] Personal Telco Project. http://www.personaltelco.net/.

[19] Pico Peering Agreement. http://www.picopeer.net/PPA-en.html.

[20] StockholmOpen.net. http://www.stockholmopen.net/.

[21] The netfilter/iptables project. http://www.netfilter.org/.

[22] The WILMA Project. http://www.wilmaproject.org/.

[23] The X-Bone. Overlay Management System. http://www.isi.edu/xbone/.

[24] Warchalking. http://en.wikipedia.org/wiki/Warchalking.

[25] WarXing. http://en.wikipedia.org/wiki/WarXing.

[26] R. Atkinson, S. Floyd, and Internet Architecture Board. IAB Concerns and Recommendations Regarding Internet Research and Evolution. RFC 3869 (Informational), August 2004.

[27] Roberto Battiti, Renato Lo Cigno, Mikalai Sabel, Fredrik Orava, and Björn Pehrson. Wireless LANs: From WarChalking to Open Access Networks. *MONET*, 10(3):275–287, 2005.

[28] Caslon Analytics. Warchalking and Wardriving. http://www.caslon.com.au/warchalknote.htm.

[29] David D. Clark, Craig Partridge, Robert T. Braden, Bruce Davie, Sally Floyd, Van Jacobson, Dina Katabi, Greg Minshall, K. K. Ramakrishnan, Timothy Roscoe, Ion Stoica, John Wroclawski, and Lixia Zhang. Making the world (of communications) a different place. *SIGCOMM Comput. Commun. Rev.*, 35(3):91–96, 2005.

[30] David D. Clark, John Wroclawski, Karen R. Sollins, and Robert Braden. Tussle in Cyberspace: Defining Tomorrow’s Internet. In *SIGCOMM ’02: Proceedings of the 2002 conference on Applications, technologies, architectures, and protocols for computer communications*, pages 347–356. ACM Press, 2002.

[31] Hans Eriksson. Mbone: the multicast backbone. *Commun. ACM*, 37(8):54–60, 1994.

[32] L. Peterson et al (eds.). Overcoming Barriers to Disruptive Innovation in Networking. *NSF Workshop Report*, January 2005.

[33] Bob Frankston. It's About Connectivity Not The Internet!, 24 February 2004. http://www.circleid.com/article/499_0_1_0/.

[34] Richard J. Gibbens, Frank P. Kelly, and Stephen R. E. Turner. Dynamic routing in multiparented networks. *IEEE/ACM Trans. Netw.*, 1(2):261–270, 1993.

[35] Mark Handley and Adam Greenhalgh. XORP : Breaking the Mould in Router Software . In *Proc. London Communications Symposium 2004 (LCS 2004)*, 13–14 September 2004.

[36] Felipe Huici, Adam Greenhalgh, Saleem Bhatti, Mark Handley, and Others. HEN Heterogeneous Experimental Network. In *Multi-Service Networks (MSN 2005) Workshop*, 8–9 July 2005.

[37] Information Society Technologies. RUNES — Reconfigurable Ubiquitous Networked Embedded Systems. http://www.ist-runes.org/.

[38] Ratul Mahajan, David Wetherall, and Tom Anderson. Understanding BGP misconfiguration. In *SIGCOMM '02: Proceedings of the 2002 conference on Applications, technologies, architectures, and protocols for computer communications*, pages 3–16, New York, NY, USA, 2002. ACM Press.

[39] UCL Department of Computer Science. Heterogeneous Experimental Network (HEN). http://hen.cs.ucl.ac.uk/.

[40] M. Papadopouli and H. Schulzrinne. Connection Sharing in an Ad Hoc Wireless Network among Collaborating Hosts. In *Proc. International Workshop on Network and Operating System Support for Digital Audio and Video (NOSSDAV)*, pages 169–185, June. 1999.

[41] Julian Priest. The State of Wireless London, 31 March 2004. http://informal.org.uk/people/julian/publications/the_state_of_wireless_london/.

[42] NRNS Information Technology Professionals. Defence R&D Canada Projects. Dynamic VPN Controller (DVC). http://www.nrns.ca/DRDC.htm.

[43] Y. Rekhter and T. Li. A Border Gateway Protocol 4 (BGP-4). RFC 1771 (Draft Standard), March 1995.

[44] Pablo Rodriguez, Rajiv Chakravorty, Julian Chesterfield, Ian Pratt, and Suman Banerjee. MAR: a commuter router infrastructure for the mobile internet. In *MobiSYS '04: Proceedings of the 2nd international conference on Mobile systems, applications, and services*, pages 217–230. ACM Press, 2004.

[45] Ricardo Sanchez, Joseph Evans, and Gary Minden. Networking on the Battlefield: Challenges in Highly Dynamic Multi-hop Wireless Networks. In *IEEE Military Communications Conference*, Atlantic City, NJ, Oct. 1999.

[46] Fred B. Schneider and Mike Rodd. International Review of UK Research in Computer Science, November. 2001.

[47] Computer Science and Telecommunications Board Committee on Research Horizons in Networks. *Looking Over the Fence at Networks: A Neighbor's View of Network Research*. The National Academies Press, National Academy of Sciences, Washington, D.C., USA, 2001.

[48] Joe Touch. Dynamic internet overlay deployment and management using the x-bone. *Comput. Networks*, 36(2-3):117–135, 2001.

[49] Eben Upton, Marc Liberatore, James Scott, Brian Levine, Jon Crowcroft, and Christophe Diot. HAGGLE: Opportunistic Communication in the Presence of Intermittent Connectivity, 26 August 2004.

[50] Petros Zerfos, Gary Zhong, Jerry Cheng, Haiyun Luo, Songwu Lu, and Jefferey Jia-Ru Li. DIRAC: a software-based wireless router system. In *MobiCom '03: Proceedings of the 9th annual international conference on Mobile computing and networking*, pages 230–244. ACM Press, 2003.